[redacted]

Oryzon Genomics S. A.

CL07-ORY-2001

A double blind, randomized, placebo-controlled, adaptive 14-week Phase IIb trial to evaluate the efficacy and safety of vafidemstat in an adult borderline personality disorder (BPD) population (PORTICO)

Statistical Analysis Plan

Version: 1.4

SPONSOR SIGNATURE PAGE

Approved by: _____ _____

[redacted]
Chief Medical Officer
CNS Clinical Development
Oryzon Genomics, S. A.

Date

A double blind, randomized, placebo-controlled, adaptive 14-week Phase IIb trial to evaluate the efficacy and safety of vafidemstat in an adult borderline personality disorder (BPD) population (PORTICO)

By signing below, all parties accept that the analysis methods and data presentations are acceptable and that this document is final.

[redacted] **Author:**

[redacted] Date

Statistical Consultant
Senior Director of Statistics
[redacted]

[redacted] **Reviewer/Approval:**

[redacted] Date

Statistical Consultant
[redacted]

TABLE OF CONTENTS

REVISION HISTORY

Version No.	Effective Date	Summary of Change(s)	Author, Institution
1.0		New document	[redacted]
1.1	19 May 2022	Update document per FDA feedback	[redacted]
1.2	Date of last signature	Update document per FDA feedback	[redacted]
1.3	Date of last signature	Update document per FDA feedback	[redacted]
1.4	Date of last signature	Update document per FDA feedback	[redacted]

LIST OF ABBREVIATIONS

Abbreviation / Acronym	Definition / Expansion
A/A	Agitation/Aggression
AAPI-CR	Agitation-Aggression Psychiatric Inventory-Clinician Report
ADME	Absorption, Distribution, Metabolism, and Excretion
AE	Adverse Event
BAC	Brief Assessment of Cognition
BDI-II	Beck Depression Inventory – II
BEST	Borderline Evaluation of Severity over Time
BMI	Body Mass Index
BPD	Borderline Personality Disorder
BPDCL	Borderline Personality Disorder Checklist
CGI-S	Clinical Global Impression of Severity
CI	Confidence Interval
CP	Conditional Power
CSP	Clinical Study Protocol
C-SSRS	Columbia – Suicide Severity Rating Scale
DMC	Data Monitoring Committee
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
EoS	End of Study
FAS	Full Analysis Set
IMP	Investigational Medicinal Product
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
MINI	Mini-International Neuropsychiatric Interview
PBMCs	Peripheral blood mononuclear cells
PD	Pharmacodynamics
PK	Pharmacokinetics
PPS	Per Protocol Set
PT	Preferred Term
SAE	Serious Adverse Event

Abbreviation / Acronym	Definition / Expansion
SAF	Safety Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
SOP	Standard Operating Procedure
STAI	State-Trait Anxiety Inventory
STAXI-2	State-Trait Anger Expression Inventory 2
STEAE	Serious TEAE
SUSAR	Suspected Unexpected Serious Adverse Reaction
TE	Target Engagement
TEAE	Treatment Emergent Adverse Events
WHO	World Health Organization

1 INTRODUCTION

This document reveals the masked information in the blinded study protocol and may not be made available to Investigators and study-related staff in direct contact with the investigators (e.g., study monitors/CRAs) until general study unblinding.

The analyses described in this SAP are based upon the following study documents:

- Study Protocol, Version 5.0 (March 31, 2023)
- Clinical Study Protocol Addendum – Unmasked Information Version 4.0 (July 23, 2021)
- electronic Case Report Form (eCRF), Study Version 5.0 (June 7, 2022)

This SAP is written to be compliant with the ICH E7 guidelines.

2 STUDY OBJECTIVES

2.1 Primary Objectives

- To investigate the efficacy of vafidemstat in the treatment of agitation and aggression in adult BPD patients
- To investigate the efficacy of vafidemstat in the treatment of adult BPD patients

2.2 Secondary Objectives

- To investigate the effect of vafidemstat in reducing the severity of BPD symptoms in adult patients
- To evaluate the safety of vafidemstat in adult BPD patients

2.3 Exploratory Objectives

- To explore the impact of vafidemstat on functional impairment in adult BPD patients
- To explore the impact of vafidemstat on cognition in adult BPD patients
- To evaluate vafidemstat plasma concentrations (pharmacokinetics) as well as vafidemstat's LSD1-target engagement in peripheral blood mononuclear cells (PBMCs) (pharmacodynamics) after treatment

3 INVESTIGATIONAL PLAN

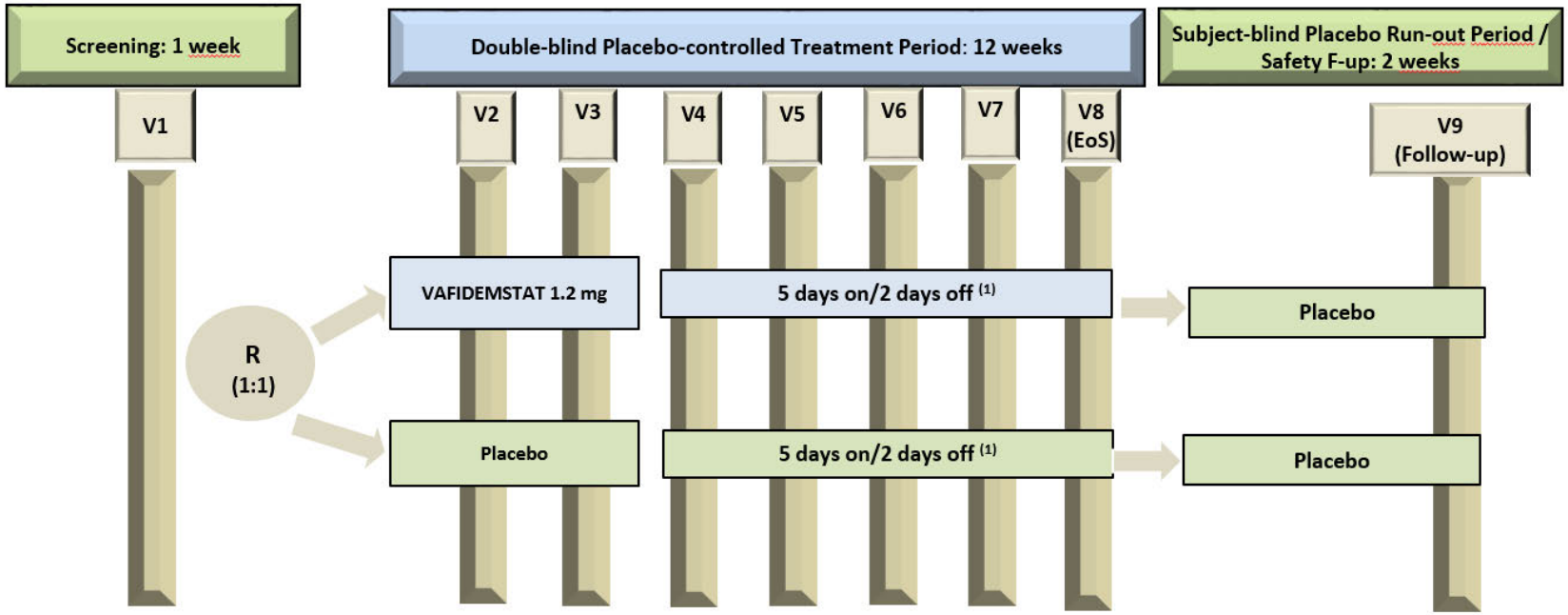
3.1 Overall Study Design and Plan

PORTICO is a double blind, randomized, placebo-controlled, adaptive 14-weeks Phase IIb trial to evaluate the efficacy and safety of vafidemstat in an adult borderline personality disorder (BPD) population.

The study will be an international, multicenter clinical trial conducted in approximately 20 study sites in Europe and the US.

Up to 188 participants will be enrolled and randomized in a 1:1 ratio (94 subjects per arm) to active treatment with vafidemstat (1.2 mg) or placebo to yield an expected 150 study subjects that complete the trial through week 14 (assuming 20% drop out). An interim analysis will be performed to assess treatment response and power to detect differences at the study endpoint after 90 subjects have completed the Week 10 assessment. The sample size may be adjusted upwards based on the outcome of the interim analysis as outlined in section 4.8.

Figure 1. Overall Study Design



(1) During the 2 days off, patients will be taking placebo capsules

Note: An interim analysis will be performed after 90 subjects have completed the *Week 10* assessments.

3.2 Endpoints

The study endpoints are listed below – all information presented in *italics* is masked in the blinded study protocol.

Primary endpoints – Efficacy

- To evaluate the difference on the Clinical Global Impression-Severity focused on Agitation/Aggression (CGI-S A/A) from baseline to *Weeks 8-12*, between the active treatment and the placebo arm.
- To evaluate the difference on the Borderline Personality Disorder Checklist (BPDCL), from baseline to *Weeks 8-12*, between the active treatment and the placebo arm

Secondary endpoints – Efficacy

- To evaluate the change overtime on the CGI-S A/A
- To evaluate the change over time on the BPDCL
- To evaluate the difference on the following measures, from baseline to *Weeks 8-12*, as well as change over time, between the active treatment arm and the placebo arm:
    a) Borderline Evaluation of Severity over Time (BEST)
    b) Beck Depression Inventory – II (BDI-II)
    c) State-Trait Anger Expression Inventory 2 (STAXI-2)
    d) State-Trait Anxiety Inventory (STAI)

Secondary endpoints – Safety

- To evaluate the following safety endpoints throughout the study, from baseline to week 14, including:
    a) Number, frequency and severity of Treatment Emergent Adverse Events (TEAEs)
    b) Number, frequency and severity of Serious TEAEs
    c) Number and percentage of withdrawn subjects due to TEAEs
    d) Frequency of physical examination parameters, vital signs and ECG parameters of potential clinical concern throughout the study period
    e) Frequency of clinical laboratory parameters (hematology, including platelets, and clinical chemistry) of potential clinical concern throughout the study period
    f) Columbia – Suicide Severity Rating Scale (C-SSRS)
    g) Use of concomitant medication throughout the study period
- To evaluate the change from baseline to *Week 12* of the following safety endpoints:
    a) Physical examination, vital signs and ECG parameters
    b) Clinical laboratory parameters (hematology, including platelets, and clinical chemistry)

Exploratory endpoints

- To evaluate the difference on the following measures, from baseline to *Weeks 8-12*, as well as change over time, between the active treatment arm and the placebo arm:

a) Columbia – Suicide Severity Rating Scale (C-SSRS)

b) Number of visits to healthcare services outside of the trial (primary care, mental health services and ER visits)

c) Agitation-Aggression Psychiatric Inventory-Clinician Report (AAPI-CR)

d) Brief Assessment of Cognition (BAC) Scale

- To evaluate vafidemstat plasma concentrations (pharmacokinetics) as well as vafidemstat's LSD1-target engagement in peripheral blood mononuclear cells (PBMCs) (pharmacodynamics) after treatment, at baseline, and at weeks 4, 8, and 12

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures, and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard Pentara procedures.

4.2 General Statistical Considerations

'Baseline' is defined as the last available non-missing pre-treatment assessment. The average of screening and baseline values will be used as the baseline for effectiveness analysis. In the instance where only one value is present that value will be solely used as baseline. Change from baseline is calculated as visit value – baseline visit. 'End of Study' is defined as the last available post-treatment assessment. 'Treatment Day' will be calculated relative to the date of randomization (i.e., Treatment Day = Assessment Date - Randomization Date + 1). For safety summaries, the last pre-treatment measurement is defined as the baseline value.

When duration is calculated for an event (e.g., Adverse Events) and starting and ending date and time are both available, then, duration will be calculated as 'End datetime' – 'Start datetime' and presented in outputs in X day(s) hh:mm format. If time is missing or partially missing for either the end or the start of the event, then, duration will be calculated in days instead as 'End Date' – 'Start Date' + 1 Day.

Continuous data will be summarized in terms of the mean, standard deviation (SD), median, minimum, maximum, and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.

Percentages will be presented to one decimal place. Percentages will be calculated using 'n' as the denominator, unless otherwise stated.

Clinical rating scales may be summarized both as a continuous and as a qualitative variable, depending on the context.

Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. P-values less than 0.001 will be presented as "<0.001".

Confidence intervals will be presented to one more decimal place than the raw data and the confidence level will be 95% unless stated otherwise.

All summary statistics will be provided by randomized or actual treatment (per analysis set used, see Section 4.5) and by visit and time point where applicable.

Visit windowing will be applied for analyses using visit categories instead of actual number of days relative to dosing for each assessment. For categorical visit summaries, all visits including early termination assessments and unscheduled visits will be included with the closest scheduled post-baseline visit that includes the effectiveness or safety assessment, based on number of days since Day 0. If multiple assessments, such as an early termination visit and a regular visit both fall within the same visit window, any non-missing effectiveness assessments will be averaged, and a worst-case approach will be used for safety data.

Days will be converted to weeks by dividing by seven. Days will be converted to months by dividing by 30.417. Days will be converted to years by dividing by 365.25. All analyses will be conducted with R v3.3.1 or SAS® v9.4 or later using procedures appropriate for the particular analysis. All data collected during the study will be analyzed and reported unless stated otherwise.

4.3 Software

All report outputs will be produced using R v3.3.1 or SAS® version 9.4 or a later version in a secure and validated environment.

4.4 Study Subjects

4.4.1 Disposition of Subjects

A clear accounting of the disposition of all subjects who enter the study will be provided, from screening to study completion.

This will include the following displays:

- A summary by randomized treatment arm (and overall) showing counts and frequency for the following:
  - Subjects screened, including reasons for screen failure
  - Re-screened subjects
  - Subjects randomized
  - Subjects treated
  - Subjects who completed the study
  - Subjects who are still on treatment (for interim analysis only)
  - Subjects who discontinued the study early, including primary reason for discontinuation
  - Subjects who discontinued treatment early, including primary reason for discontinuation
- The number of subjects screened and randomized will also be displayed by country and center

These summary displays will be using all available data (all subjects). Percentages will be based on the number of subjects randomized.

In addition, the following listings will be produced:

- Informed consent (including any re-consents to updated protocol versions, if applicable)
- Randomization (including date of randomization, treatment randomized to and actual treatment, and whether the blind was broken pre-maturely)
- Treatment discontinuations, including duration of treatment and reason for discontinuation

4.4.2 Protocol Deviations

Major protocol deviations are defined as those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments. The impact of major protocol deviations on the efficacy and/or safety results will be investigated by assessing the robustness of the study results and conclusions to the choice of analysis set, both including and excluding data potentially affected by major protocol deviations.

Major protocol deviations and any action to be taken regarding the exclusion of subjects or affected data from specific analyses are defined in the project-specific Protocol Deviation Specification. However, the final decision for exclusions and inclusions will be taken during the Blinded Data Review Meeting (BDRM) before unblinding the database.

The following outputs will be produced:

- A summary of major protocol deviations by randomized treatment arm and overall, including:
  - Number and percentage of patients having at least one major protocol deviation
  - Number and percentage of patients by protocol deviation category (i.e., informed consent, missed assessments, …)
- Listing of all major protocol deviations
- Listing of all minor protocol deviations
- Listing of all inclusion/exclusion criteria violations

Note that inclusions/exclusion criteria will be included both in the specific listings for inclusion/exclusion criteria deviations and in the listing for the appropriate major/minor classification.

4.5 Analysis Sets

With regard to analysis of data, the following analysis sets are defined:

- **Safety Analysis Set (SAF):** All subjects who received at least one dose of the study IMP or Placebo. Individual data will be summarized as treated.
- **Full Analysis Set (FAS):** All patients enrolled (randomized) in the study. Patients will be analyzed as randomized.
- **Per-Protocol Set (PPS):** All subjects in the FAS without any major protocol deviations* as described in Section 10.9 of the Clinical Study Protocol (CSP) and who also satisfy both of the following constraints:
  1. Have at least completed one CGI-S A/A assessment at *Weeks 8-12*
  2. Have at least completed one BPDCL assessment at *Weeks 8-12*

  Individuals in the PPS population will be analyzed as treated.

*Important violations of eligibility criteria and other deviations from the protocol will be assessed in cooperation with the Sponsor. Important deviations from the protocol may lead to exclusion of a subject from the PPS (See Blinded data review below).

Primary efficacy analysis will be performed using the FAS. Patients will be analyzed as per randomized treatment. Sensitivity and supplementary analyses will be performed in the FAS. Efficacy will also be tested in the PPS analysis will be for descriptive purposes only.

The safety summaries and analyses will be based on SAF. The PPS may be used for additional safety analysis.

Before database lock, protocol deviation and analysis set outputs will be produced and will be sent to Oryzon for review. Analysis Set classifications will be discussed during the Blinded Data Review Meeting to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects and/or subject data from analyses will be made prior to unblinding and will be documented and approved by Oryzon.

A summary of patients (counts and frequencies) in each analysis set, including reasons for exclusion, will be produced in addition to a listing of analysis set assignments by patient, including any reasons for exclusion.

4.6 Estimands

Summary of primary estimand:

1. In study participants with adult BPD, what is the effect of treatment taken for 12 weeks relative to placebo during the double-blind phase, in the protocol-defined FAS study population as measured by the primary endpoints Clinical Global Impression-Severity focused on agitation/aggression (CGI-SS A/A) and Borderline Personality Disorder Checklist (BPDCL).
2. *Objective*

   An estimate of effectiveness which would potentially be supportive of registration. Clinically meaningful effects greater than or equal to 0.2 will be supportive of proof of concept.

3. *Estimand*

A. The population is the FAS as defined by the protocol's inclusion and exclusion criteria.

B. Effectiveness is measured using an MMRM to estimate change from baseline over 8-12 weeks in the primary endpoints (CGI-SS A/A and BPDCL).

C. The treatment evaluated is the randomized treatment as it was assigned to study participants. All types of intercurrent events are incorporated using a treatment policy strategy, which evaluates the randomized treatment as taken including missed or modified treatments, treatment discontinuation, and concurrent treatments. Deaths due to treatment are not expected during the treatment period. A hypothetical strategy will be used to account for the intercurrent event of early study termination to estimate the effect anticipated if all patients had completed the study.

D. The mean change from baseline between treatment and control arms on the primary endpoints will be estimated over visits associated with 8 -12 weeks of treatment. The study is controlled at a 5% level. The Hochberg correction will be applied. If one endpoint is below the 2.5% level or if both endpoints are below the 5% level of significance the study is declared a success. Intercurrent events that are expected to occur include early treatment discontinuation due to patient choice or due to adverse events both unrelated and related to the medication or study indication (e.g., COVID-19).

4. *Primary estimator and missing data.*

   The estimator for the primary estimand is a likelihood-based mixed model for repeated measures (MMRM). The primary analysis will include all available assessments. Individuals with no post-baseline information will have the first post-baseline visit imputed. Individuals reported to drop out for lack of efficacy will be given the baseline placebo mean. Individuals reported as dropping out for any other reason will be given the baseline group mean of their corresponding treatment.

5. *Sensitivity analyses and missing data.*

To assess the effect of averaging efficacy for visits corresponding to weeks 8-12 on the primary outcome, a sensitivity analysis will be implemented in which the primary efficacy analysis will be run for visits associated with weeks 8, 10, and 12 separately. As clinical studies are limited in the ability to schedule visits on the exact prespecified study day a continuous time model that extracts the predicted estimates on the prespecified study day will be used as a sensitivity model to assess the degree to which windowing affects the outcome. Additional sensitivity models will be run with multiple imputation within each treatment arm. The first analysis (missing at random; MAR) is meant to estimate the treatment effect if all study participants continued on their current trajectory within the treatment group by imputing the missing data within associated treatment groups. The second analysis (missing not at random; MNAR) imputation is performed for all study participants relative to the placebo group only. This analysis assumes study participants at dropout progress similarly to the placebo group. This analysis will provide a lower bound estimate of the effect. A tipping point analysis using multiple imputation will also be performed.

4.7 Determination of Sample Size

The sample size for the study was calculated assuming an effect size of 0.51 on both endpoints between active treatment and placebo. Controlling the overall type I error rate at 10% for the study, each endpoint will be tested at a two-sided alpha of 5% (equivalent to 2.5% one-sided) considering that a significantly larger decrease compared to placebo in either endpoint will be considered sufficient for study success. A power of approximately 80% for each endpoint is achieved with 62 patients completing the study in each arm, active treatment, and placebo. Accounting for a drop-out rate of 20%, 78 patents (total 156 patients) should be recruited per arm. A sample size adjustment will be considered at the interim analysis (see Section 9.6).

The original sample size adjustment above controls the overall type 1 error rate at 10%. However, to meet the requirements for proof of registration, the alpha was adjusted to control for a type 1 error rate of 5%, split evenly between the two endpoints. Assuming an effect size of 0.51 (mean 510, standard deviation 1000) a sample size of 150 completers (75 per arm) is required for 80% power at a two-sided alpha of 0.025 for each endpoint. Assuming 20% dropout would require 188 patients enrolled.

4.8 Interim Analyses

An interim analysis will be conducted after 90 patients have completed their *Week 10* evaluations (expected 108 patients enrolled after accounting for 20% dropout). Individuals that had the opportunity to complete their Week 10 visit but were lost to follow-up will also be included in the analysis, however, will not be included in the 90-patient count required for interim analysis initiation. In addition, unblinded safety outputs will be produced. After the patient count requirement for interim analysis has been reached, sufficient time will be allowed for data cleaning, data transfers (including laboratory) and medical coding.

The interim analysis will be conducted as per the main efficacy analysis described in Section 5.1.13. Sensitivity analyses will not be performed.

A sample size adjustment will be performed using the promising zone approach as described by Mehta and Pocock ([1]) which allows the sample size to be adjusted based on an interim analysis only if the data are in a promising zone, i.e., the effect is trending toward success but not strong enough to indicate it will finish successfully without increasing the sample size. If the effect is

strong (favorable zone) or too weak (unfavorable zone) the trial will be allowed to accumulate participants to the originally planned sample size (188 enrolled; 150 anticipated completers). If data are within the promising zone, the sample size will be increased in a stepwise fashion to keep the estimated power of the study between 80 and 85%. This step wise approach to increasing sample size at interim is implemented to avoid the possibility of back calculating the observed effect size from the sample size increase, although variations in standard deviation and placebo decline rates from historical expectations already make this difficult.

As Mehta and Pocock assess single outcomes, we interrogated control of the type I error rate using dual outcomes as proposed in this SAP. The code to generate the simulation and results showing that this approach effectively controls type I error in a Promising Zone simulation can be found in the attached document promising_zone_simulation.R. Images produced with this code can be found in Appendix 7.2.

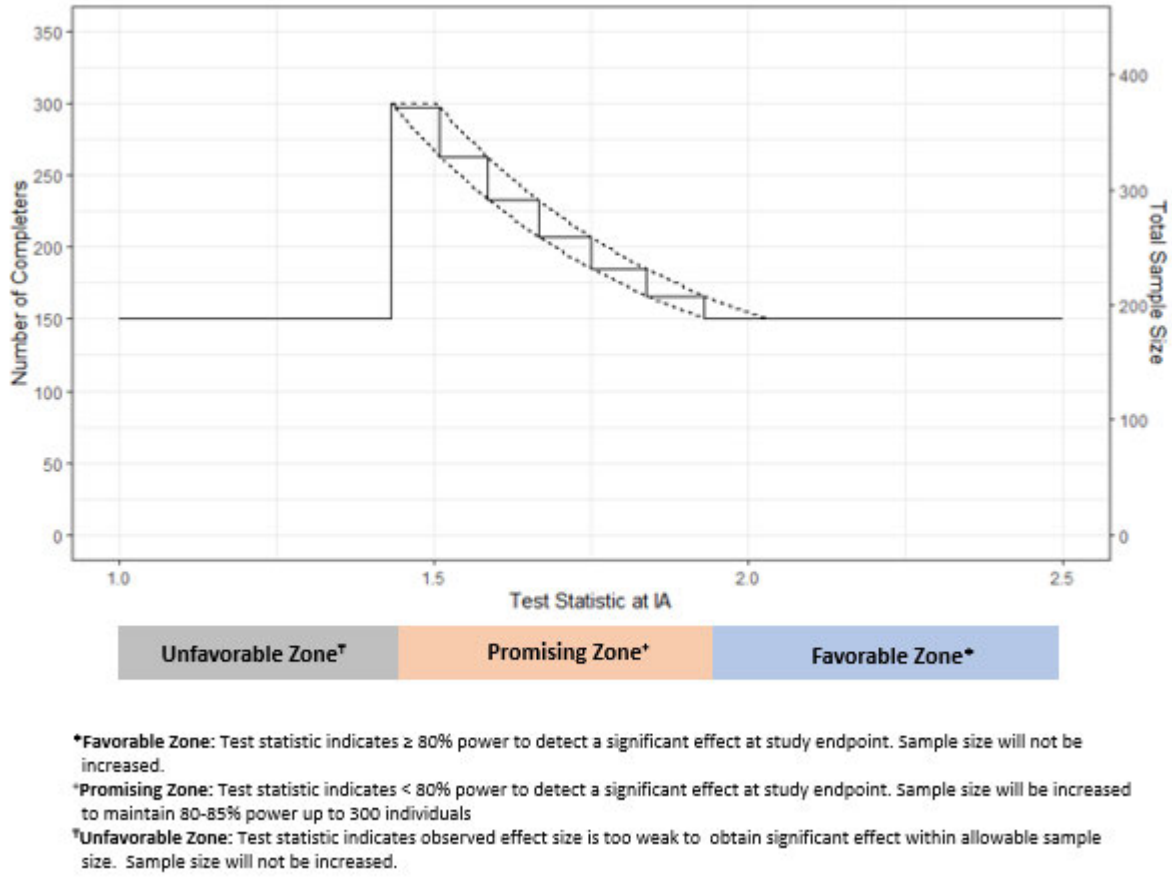


Figure 2. Promising Zone. The promising zone boundaries (orange box) and corresponding sample size increases are detailed above. The solid black line corresponds to the proposed enrolled (right; y-axis) and completer (left; y-axis) sample sizes relative to the test statistic (x-axis) observed at interim. Sample sizes were calculated using an 80-85% power threshold to provide binning for sample size increase, thus avoiding back calculation of the effect size at interim. Unfavorable and favorable zones are shaded in grey and blue, respectively.

The conditional power at interim, or probability of study success, for detecting the difference in primary outcome between the two groups in the final analysis, will be estimated using the accumulated data. The proposed starting sample size of 90 study participants results in 85% power for both primary outcomes (assuming a Cohen's D of 0.51). We have elected to begin the study

with the expected sample size estimate and adjust upward if improvement in the treated cohort does not resemble what was observed in the previous studies.

The promising zone will increase the sample size to maintain 80% to 85% power to detect a significant effect, i.e. if the power goes below 80% the sample size will be increased to achieve power between 80% and 85% (Table 1; Figure 2). Development of the 80-85% thresholds also maintain blind of the interim results. If at interim the test statistic is in the favorable zone (≥ 1.944) the study will proceed as originally described (150 completers). If at interim the test statistic falls within the promising zone (test statistic: < 1.944 ≥ 1.412), sample size adjustment will follow the test statistic ranges outlined in Table 1 and Figure 1. If the test statistic is less than 1.412 (40% conditional power at interim) the sample size will not be increased (remain at 150 completers). The lower bound of the promising zone (effect size 0.23) is set by sample size constraints, however clinically meaningful effects have been observed at an effect size as low as 0.2. The study will be stopped for futility if the test statistic falls below a 50% probability (conditional power) of observing a result favoring *placebo*. This corresponds to a test statistic below -1.987 (effect size equal to -0.4177).

At interim analysis two discontinued individuals reached visit 6, and were included in the IA analysis. The table below was adjusted for 92 individuals worth of information, as opposed to the 90 individuals mentioned above.

Table 1. Promising Zone Thresholds

	Test Statistic at Interim	Completers	Sample Size (total Enrollment)
Futility	$-\infty < Z \leq -1.987$	N at time of futility declaration	--
Unfavorable Zone	$-1.987 < Z < 1.412$	150	188
Promising Zone	$1.412 \leq z < 1.555$	300	376
	$1.555 \leq z < 1.649$	265	332
	$1.649 \leq z < 1.745$	231	290
	$1.745 \leq z < 1.845$	201	252
	$1.845 \leq z < 1.944$	174	218
Favorable Zone	$1.944 \leq z < \infty$	150	188

Sample size calculations will be generated as described below. The more conservative sample size calculation between the two primary efficacy endpoints will be used.

The following formulas will be used:

$$CP(z_1, \tilde{n}_2) = 1 - \Phi\left(\frac{z_\alpha\sqrt{n_2} - z_1\sqrt{n_1}}{\sqrt{\tilde{n}_2}} - \frac{z_1\sqrt{\tilde{n}_2}}{\sqrt{n_1}}\right)$$

$$\tilde{n}_2' = \left(\frac{n_1}{z_1^2}\right)\left(\frac{z_\alpha\sqrt{n_2} - z_1\sqrt{n_1}}{\sqrt{n_2 - n_1}} + z_\beta\right)^2$$

Here:

- z_1 is the Wald-statistic observed at the interim analysis,

- α is the *one-sided* significance level of the test (here 2.5%),
- z_α is the corresponding critical value of the normal distribution,
- n_2 is the total number of patients originally planned,
- $\tilde{n}_2$ is the incremental number of patients $n_2 - n_1$,
- $\tilde{n}_2'$ is the proposed new incremental number of patients,
- β is 1 - statistical power, and
- Φ is the cumulative density function of the normal distribution.

Note: $\tilde{n}_2'$ will only be calculated if CP is in the promising zone. Both n_2 and the adapted sample size $n_2' = n1 + \tilde{n}_2'$ are adjusted for the expected drop-out rate, i.e., the n_2 is considered to be 150 and $n_2'/0.8$ patients (rounded up to an even number) will be recruited after adaptation of the sample size.

If the CP for both endpoints is below 40%, the study will be continued without modifications.

Details for the safety deliverables are described in Section 5.2.11.

All analyses at interim will be conducted by a separate unblinded statistical team.

The decision tree for promising zone can be found in appendix 7.3.

4.9 Demographic and Other Baseline Characteristics

Demographics will be based on the SAF. The following summaries will be produced:

- A summary of demographic variables (e.g., age, sex, race, height, and weight) by treatment and overall
- A summary of disease background factors, including whether patient is in psychotherapy at the start of the study
- A summary of the medical history, by System Organ Class and Preferred Term

Demographics and baseline characteristics will also be listed by patient.

The answers to the Mini-International Neuropsychiatric Interview (MINI) will be listed, and the primary diagnosis will be summarized as part of the disease background factors table.

4.10 Concomitant Medication

Medication start and stop dates will be compared to the date of first dose of study medication to allow medications to be classified as either Prior only, both Prior and Concomitant, or Concomitant only. Medications starting after the completion/withdrawal date will be listed but will not be classified or summarized.

Medications that start and stop prior to the date of first dose of study medication will be classified as Prior only. If a medication starts before the date of first dose of study medication and is either ongoing or stops on or after the date of first dose of study medication, then the medication will be classified as both Prior and Concomitant. Medications will be classified as Concomitant only if they have a start date on or after the date of first dose of study medication.

If medication start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of first dose of study medication. Medications will be assumed to be Concomitant only, unless there is clear evidence (through comparison of partial dates) to suggest that the medication started prior to the first dose of study medication. If there is clear evidence to

suggest that the medication started prior to the first dose of study medication, the medication will be assumed to be both Prior and Concomitant, unless there is clear evidence to suggest that the medication stopped prior to the first dose of study medication. If there is clear evidence to suggest that the medication stopped prior to the first dose of study medication, the medication will be assumed to be Prior only.

Prior and Concomitant medications will be coded according to WHO Drug Dictionary.

Summaries of both Prior and Concomitant medication will be provided for the SAF analysis set by actual treatment arm, displaying the number of percentage of subjects being treated with each type of medication/therapy classified according to ATC level 2 (therapeutic main group) and ATC level 3 (pharmacological subgroup). An additional summary will be created for only prohibited medication (as per Table 2 of the CSP).

A similar summary will be provided for Prior medication.

Medication will also be listed, including classifications, start and end dates, indication, frequency and dose as well as other available variables.

4.11 Treatment Compliance

Compliance (%) is calculated in the eCRF by visit as actual capsules [dispensed – returned] / planned capsules.

Overall compliance will be derived as the mean compliance across all visits where compliance is calculated in the eCRF, i.e., visits for which drugs were dispensed and returned.

The eCRF calculation may not properly account for all deviations from the dispensation and dosing schedule, and will not be reported in output tables. Possible examples for this include delayed return of dispensed capsules or taking medication in the morning instead of in the clinic for scheduled study visits. Therefore, a derived subject-level compliance (%) will also be calculated in the analysis datasets and reported for each dispensation period as follows:

$$\text{Compliance(period)} = \frac{\text{Capsules Dispensed} - \text{Capsules Returned}}{\min(\text{Date Return, Date next Dispensation}) - \text{Date Dispensation}}$$

A mean subject-level compliance will be calculated per patient as the average compliance of periods where the dates required to calculate the period specific compliance are available. Subject-level compliance will be not be calculated for individuals with missing information.

Protocol deviation records will be used to identify cases where a subject took a capsule in the morning instead of on-site and those cases will be flagged in listings and the number and frequency of patients with such cases will be summarized in the exposure listing.

Compliance will be further investigated by running a Chi-Square Test of the observed vs the expected Ctrough and/or Target Engagement values. This analysis will be planned and described separately from this SAP. Exposure information from the eCRF will be listed, including date dispensed, first intake, amount caps, date of return, compliance and how many caps were returned.

In addition, the number and frequency of patients whose target engagement levels are below 50% [ratio, occupied/total] will be summarized by visit as a likely indicator of non-compliance. This will be used for the primary efficacy analysis. All available data will be included, but missing values will not be imputed so that patients missing *Week 12* data will not contribute directly to the estimation of the main contrast of interest. Alternative options will be explored for sensitivity analysis, with various imputation methods including multiple imputation.

4.12 Multiple Comparisons/Multiplicity

There are two primary endpoints for efficacy analysis: CGI S A/A and BPDCL, a significant difference in either one is considered sufficient for study success. The overall significance level of the study is thus strictly controlled at a level of 5%. The Hochberg correction will be applied. If one endpoint is below the 2.5% level or if both endpoints are below the 5% level of significance the study is declared a success.

5 Efficacy Evaluation

5.1.1 Primary Efficacy Variable – CGI-S A/A

All available data will be included in the analysis.

This section covers the primary analyses of CGI-S A/A and BDPCL as well as the secondary endpoints of change over time for the same assessment.

The primary endpoints for the assessment of efficacy are the change from baseline to *Weeks 8-12* in CGI-S A/A and BDPCL. Results will be summarized by treatment and visit in terms of absolute values and changes from baseline. Besides summarizing the numerical values, the number and frequency of patients at each value on the scale will be displayed (from 1 to 7).

Boxplots of changes in CGI-S A/A will be presented by treatment arm versus each visit. In addition, individual and mean absolute values and changes from baseline will be plotted versus week.

A shift table will be produced for CGI-S A/A displaying the frequencies and proportions for the following

- At *Weeks 8-12 separately as well as an combining the model estimates of the endpoint values over 8-12 weeks*:
  - Any decrease
  - Any increase
  - A decrease of at least 3 points
  - An increase by at least 3 points
- Any post-baseline value higher than the baseline
- Any post-baseline value lower than the baseline

The effect of treatment in terms of the change from Baseline to *Weeks 8-12* in CGI-S A/A and BDPCL will be analyzed using a likelihood-based mixed effects model for repeated measures (MMRM) with changes from baseline to each post-baseline visit as the dependent variable and the following factors:

Fixed Factors:

- Site (SITEGR1)
- Visit (AVISITN)
- Treatment Arm (TRT01P)
- Interaction between Visit and Treatment Arm
- Baseline value of the response variable*
- Psychotherapy at baseline (yes/no) (PSYBLYN)*
- Psychotherapy by baseline interaction*

*Interactions with time (visit) will be included for all covariates listed above.

The random effect of subject will be accounted for via the within-subject error correlation structure. The preferred covariance structure for the model will be unstructured (UN in SAS), If this analysis fails to converge, the model will be simplified first to Heterogeneous Toeplitz (TOEPH), then first order autoregressive with heterogeneous variance (ARH1), then compound symmetry with heterogeneous variance (CSH), followed by compound symmetry (CS). The first (co)variance structure yielding convergence will be used as the primary analysis. If a structured covariance matrix is used the sandwich estimator will be implemented to account for potential correlation matrix misspecification. The degrees of freedom will be calculated using the general Kenward-Roger approximation.

The SAS code for the primary analysis is as follows:

```
PROC MIXED data=adqs;
CLASS TRT01P(REF='Placebo') AVISITN(REF='3') SITEGR1(REF='01') PSYBLFL USUBJID;
MODEL CHG= SITEGR1 AVISITN TRT01P AVISITN*TRT01P STDBASE STDBASE*AVISITN PSYBLFL
STDBASE*PSYBLFL / s SOLUTION CL DDFM = KenwardRoger;
REPEATED AVISITN/ SUBJECT= USUBJID TYPE=UN;
LSMEANS TRT01P*AVISITN /at STDBASE=0 PDIFF CL;
ESTIMATE \see below\\
Run;
```

Estimates statements averaging visits 6-8 are as follows:

Vafidemstat LSMEANS

```
ESTIMATE 'TRT01P{Vafidemstat} AVISITN{AVG 6-8}'
        INTERCEPT 1
        AVISITN 0 0 0.33333 0.33333 0.33333 0
        TRT01P 1 0
        AVISITN*TRT01P 0 0 0.33333 0.33333 0.33333 0 0 0 0 0 0 0/ e cl;
```

Placebo LSMEANS

```
ESTIMATE 'TRT01P{Placebo} AVISITN{AVG 6-8}'
        INTERCEPT 1
        AVISITN 0 0 0.33333 0.33333 0.33333 0
        TRT01P 0 1
        AVISITN*TRT01P 0 0 0 0 0 0 0 0 0 0.33333 0.33333 0.33333 0 / e cl;
```

LSDiff (Vafidemstat – Placebo)

```
ESTIMATE 'TRT01P{Vafidemstat} AVISITN{AVG 6-8} _TRT01P{Placebo} _AVISITN{AVG 6-8}'
        TRT01P 1 -1
        AVISITN*TRT01P 0 0 0.33333 0.33333 0.33333 0 0 0 -0.33333 -0.33333 -0.33333 0 / e cl;
```

Sites with less than two study participants per arm will be pooled. To pool sites: sites with less than two study participants per arm will be ranked within country by the number of participants within the placebo arm (higher numbers are ranked higher). Ranked sites will be pooled starting by pooling the highest ranks until the study participant requirement per arm is met. If more sites are present after the pooling criterion is met the remaining sites are pooled to meet the criterion. In the event that remaining sites do not meet the criterion they will be combined with the previously pooled group. In the event of a tie for ranking the sites, the site with the largest number in the active arm will be used. If the tie persists the sites will be ordered alphabetically or numerically (A and higher number result in a higher rank). Due to the small sample size the interim analysis will be ranked across all sites for pooling as opposed to within country.

Pooling at interim will not include country due to the small number of individuals.

The primary treatment contrast will be based on the LS means from the treatment by visit interaction term. A contrast statement will be written such that model estimates from assessments collected between weeks 8-12 are averaged to compare drug vs placebo.

The test will be a two-sided test of the null hypothesis of no difference between active treatment and placebo at *Weeks 8-12*.

A listing by subject of the CGI-S A/A scores including changes from baseline will be provided.

Individuals with no post-baseline information will have the first post-baseline visit imputed. Individuals reported to drop out for lack of efficacy will be given the baseline placebo mean. Individuals reported as dropping out for any other reason will be given the baseline group mean of their corresponding treatment.
Reasons not associated with lack of efficacy include but are not limited to the following:

- Participant Decision Due to COVID
- Lost to Follow-up
- Adverse Event
- Protocol Deviation
- Physician Decision
- Non-Compliance
- Blood Value During Randomization Visit Was Exclusionary
- Withdrawal by Subject

For the primary efficacy analyses all other missing data will be accounted for using the MMRM estimates.

The analysis of Total BPDCL will not include a categorical summary table.

5.1.1.1 Sensitivity Analysis

Sensitivity analysis for the CGI-S A/A will focus on the following areas:

Estimate over 8-12 week visits

To assess the degree to which estimating across visits represents the overall effect observed, the primary analysis will be repeated for visits 8, 10, and 12 separately. The combined estimate approach will be revisited if the individual visit analyses are not directionally consistent or if this approach is determined to change the direction of the outcome effect, i.e. one visit is driving the overall effect in the primary analysis (ex: one visit significantly favors treatment, while the other two show no effect or favor placebo.

Analysis Population

The primary analysis will be repeated using the PPS for descriptive purposes only.

Missing Values: Multiple Imputation

To conduct further sensitivity analysis intercurrent events like discontinuation of treatment or patient withdrawal, will be imputed. The PROC MI procedure in SAS will be used to do predictive mean matching using linear regression models ([2]). Imputation will be performed after window mapping. The data at each week will be predicted only using data from the previous weeks. The following SAS model will be used:

```
proc mi data=manual nimpute=30 out=MI1 seed= 123456;
  class trt;
  var trt baseline week2 week4 week6 week8 week10 week12;
  fcs regpmm (week2 = baseline trt baseline*trt /details);
  fcs regpmm (week4 = baseline trt baseline*trt week2 week2*trt /details);
  fcs regpmm (week6 = baseline trt baseline*trt week2 week2*trt week4 week4*trt
/details);
  fcs regpmm (week8 = baseline trt baseline*trt week2 week2*trt week4 week4*trt
week6 week6*trt /details);
  fcs regpmm (week10 = baseline trt baseline*trt week2 week2*trt week4 week4*trt
week6 week6*trt week8 week8*trt /details);
  fcs regpmm (week12 = baseline trt baseline*trt week2 week2*trt week4 week4*trt
week6 week6*trt week8 week8*trt week10 week10*trt /details);
run;
```

Here, baseline is the baseline measurement, and `week x` is the result of the measurement for Week x. `Trt` is an indicator variable for the received treatment.

Imputed values that fall outside of the possible range of the test variable will be set to the min or max possible score for the test before imputation.

The following simplifications will be attempted if the model fails to converge for any reason:

1. Only the results from the directly preceding visit and baseline will be used in each of the models (i.e., Week 10 values will be predicted using Week 8 and baseline measurements)
2. Only baseline and baseline*trt will be used in each prediction model

Change from baseline to *Weeks 8- 12* will be calculated for each of the 30 simulated datasets. The main analysis model described in 5.1.1 will be run for each simulated dataset and the values will be aggregated using PROC MIANALYZE. In addition to the estimates for the contrast between treatments at *Weeks 8- 12* (including confidence intervals and p-values), the composition of the total variance in parameter estimation will be broken down by variability between datasets and variability within datasets.

This analysis above will be performed using the entire dataset for imputation across treatment as well as imputation based on the placebo group only.

A tipping point analysis as implemented in Ratitch et al. [3] will also be run. This analysis finds a tipping point across a spectrum of assumptions underlying the missingness of the data (from less conservative to more conservative) in which the conclusion of the study changes from favoring the treatment group to being unfavorable.

Impact of Covariables

The impact of additional covariables will be investigated by including them in the main model, including interactions with Visit and Treatment Arm, as well as the triple interaction between the covariable, Visit and Treatment arm.

The following covariables will be considered:

- Country or center (if number of patients per center allows)
- Psychotherapy at baseline (the main effect is already part of the standard model, so only the interactions need to be added for sensitivity)
- Sex

Other demographic variables may be investigated for sensitivity analysis if there is a reason to suspect an imbalance between treatment arms.

These covariables will be tested by including only one covariable (and interactions) at a time, not all at once. F-tests will be performed for the main effect as well as interactions. In case of highly significant interactions further analysis will be considered. This sensitivity analysis will be performed using the FAS.

Model Assumptions

As described in Section 5.1.1 the assumptions of the mixed model will be investigated visually by the statistician. Regardless of the outcome of this investigation, at least the following sensitivity analyses will be performed:

- A Welch's t-test of the changes from baseline at *Week 12* between treatments without any adjustment for covariables

This sensitivity analysis will be performed using the PPS.

5.1.2 Primary Efficacy Variable - BPDCL

The analysis of Total BPDCL will be the same as described in Section 5.1.1 for CGI-S A/A, except that no categorical summary table will be provided.

5.1.3 Secondary Efficacy Variables

For the secondary efficacy variables, Borderline Evaluation of Severity over Time (BEST), Beck Depression Inventory – II (BDI-II), State-Trait Anxiety Inventory (STAI) and State-Trait Anger Expression Inventory 2 (STAXI-2).

The following will be the primary variables used for statistical analysis and/or summary for each of the tests:

- BDI-II: Total Score
- BEST: Score (sum) for each of the individual 3 sub-categories (thoughts and feelings, behaviors [negative], behaviors [positive]), Total Score
- STAI: S-Anxiety and T-Anxiety Raw Scores
- STAXI-2: State Anger Scale Raw Score, Trait Anger Scale Raw Score and AX Index Raw Score

Summary statistics for each of those items will be presented by treatment and visit using the FAS. Mean scores (± standard deviation) will be presented by treatment.

The same MMRM as described in Section 5.1.1 will be calculated for secondary efficacy variables which are scheduled to be collected at *Weeks 8-12* (e.g., for STAI only S-Anxiety is collected at each visit), except that there will not be a formal significance test for the difference between treatments at *Week 12* and that no sensitivity analysis is planned for these endpoints. The differences from baseline at each week will be reported, including a 95% confidence interval.

The answers for each test item will be listed by subject and scores will be presented in a separate listing.

The results for all other efficacy assessments mentioned in Section 7.2 of the protocol will be listed and summarized by treatment/Visit.

5.2 Safety Evaluation

All safety summaries and analyses will be based upon the SAF as defined in Section 4.5.

5.2.1 Extent of Exposure

The total dose taken will be estimated using compliance data, by taking the difference of dispensed vs returned caps. Exposure will not be calculated for patients with an unknown number of returned caps.

Target Engagement and pharmacokinetic (Ctrough) values will also be summarized by treatment and visit. A listing for Target Engagement and Ctrough values will be presented.

5.2.2 Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1 or higher.

Treatment-emergent adverse events will be tabulated and are defined as those adverse events that either start or worsen in severity on or after the date/time of first dose of study treatment.

Where dates are missing or partially missing, adverse events will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the adverse event started prior to the first dose of study treatment.

For summary tables, AEs with “Certain”, “Probable/Likely” and “Possible” relationship to the study drug, as well as all AEs judged as unassessable/unclassifiable, will be considered related. In listings the more granular classification will be used.

The following AEs are considered to be of special interest:

- Investigations (SOC):
    - Platelet count decreased (PT)
    - Neutrophil count decrease (PT)

Only platelet and neutrophils count decrease with hematological impact (i.e., at a given study visit, a decrease in the platelet or neutrophil count *equal to or more than a 30% from baseline´s count*) are to be reported as an AE unless, in the opinion of the investigator, the decrease is non-clinically significant.

An overall overview of TEAEs will be presented, including number of patients and frequency affected by different types of AEs as well as the number of events for each category by treatment:

- Any TEAEs
  - TEAEs related to study drug
  - TEAEs leading to study discontinuation
  - TEAEs by severity (mild, moderate severe)
  - TEAEs by outcome
- Any Serious TEAEs (STEAEs):
  - STEAEs related to study drug
  - STEAEs leading to study discontinuation
  - STEAEs by severity (mild, moderate severe)
  - STEAEs by outcome
- Any TEAEs of special interest
  - Platelet count decreased
  - Neutrophil count decreased
- Any COVID-19 TEAEs

The total number of subjects with at least one TEAE and the total number of TEAEs will be presented overall and by treatment arm and tabulated by MedDRA SOC and PT. In addition, AEs will be tabulated by worst severity and worst relationship to treatment.

The same tabulations will be provided for TEAEs, TEAEs of special interest, TEAEs leading to study drug discontinuation, serious TEAEs (STEAEs) and STEAEs related to study medication.

Listings of subjects with STEAEs, deaths and STEAEs leading to discontinuation and individual narratives for these cases will be provided.

Adverse event summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the vafidemstat treatment, and then similarly by decreasing frequency in the placebo treatment, and then alphabetically for SOC, and PT within SOC.

A listing by subject of all adverse events (including non-treatment-emergent events) will be provided. This listing will be presented by treatment and will include the following: center, subject identifier, age, sex, race, adverse event (SOC, PT, and verbatim term), date of onset, date of resolution, duration, severity, seriousness, action taken, outcome and causality.

If the start date/time of an adverse event (AE) is partially or completely missing, the date/time will be compared as far as possible with the date/time of the start of administration of study treatment. The AE will be assumed to be treatment-emergent if it cannot be definitively shown that the AE did not occur or worsen during the treatment-emergent period (worst-case approach). The following general rules will be used:

- If the start day is missing but the start month and year are complete, an AE will only be excluded as being treatment-emergent if the start month/year is before the month/year of first date of treatment or if the stop date/time is before last date of treatment.

- If the start day and month are missing but the start year is complete, an AE will only be excluded as being treatment-emergent if start year is before the year of first date of treatment or if the stop date is before first date of treatment.

- If the start date is completely missing, an AE will be considered treatment-emergent unless the stop date/time is before first date of treatment.

For other outcomes (e.g., date of vital signs collection) missing day will be filled in with the middle day of the month (e.g., day 15). If month and day are missing, they will be filled in with July 3rd (the middle day of the year).

5.2.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events

The following listings will be produced:

- A by-subject listing of all deaths that occurred during the study
- A by-subject listing of all serious adverse events
- A by-subject listing of all adverse events leading to discontinuation of study treatment
- A by-subject listing of all other adverse events of special interest

5.2.4 Clinical Laboratory Evaluation

For summaries by visit, the last non-missing assessment (including repeat assessments) recorded at each visit will be summarized. Unscheduled and unplanned assessments will be listed, but not included in by-visit summaries.

For across visit summaries (e.g., maximum post-baseline value), scheduled, unscheduled and repeat assessments will be considered.

A subject will be defined as having a treatment-emergent laboratory abnormality if the subject had a normal baseline value and an abnormal post-treatment start value.

The following summaries will be provided for each laboratory category:

- A summary of each laboratory parameters (absolute values and change from baseline) by treatment and visit
- A summary of the number and percentage of patients experiencing treatment emergent laboratory abnormalities by parameter, visit (and overall, throughout the study)
- A shift from baseline table by visit

Laboratory values for hematology, serum biochemistry, serological parameters and urine parameters will each be listed by subject and study time point including changes from baseline (with the exception of serological and urine parameters). The baseline for the laboratory values will be the last results obtained prior to treatment (Day 1 according to schedule)

All values outside the clinical reference ranges will be flagged in the data listings. The abnormal values will be flagged with 'L' for values below the lower limit of the clinical reference range and 'H' for values above the upper limit of the clinical reference range and included in the listings. The Investigator will assess whether the values outside the clinical reference range are clinically significant. Clinically significant laboratory values will be recorded by the Investigator as AEs.

5.2.5 Vital Signs

Vital signs results will be listed by patient and visit, including changes from baseline and assessments as abnormal/normal, and if abnormal, as clinically significant or not clinically significant.

Descriptive statistics will be provided for each vital sign parameter by visit/time point and treatment, including changes from baseline.

5.2.6 Physical Examinations

The results of the physical examination will be listed by subject.

5.2.7 Pregnancy Tests

The results of the pregnancy tests will be listed by subject.

5.2.8 Urine Drug Screen Results

The results of the urine drug screen will be listed by subject.

5.2.9 Electrocardiograms

ECG parameters will be listed by patient and visit, including changes from baseline and overall assessment as abnormal/normal, and if abnormal, as clinically significant or not clinically significant.

Descriptive statistics will be provided for each parameter by visit/time point and treatment, including changes from baseline. Categorical summaries of ECG findings will be presented in an additional table.

5.2.10 Columbia-Suicide Severity Rating Scale

The C-SSRS results will be listed.

The following items will be summarized by treatment and visit:

- Number and percentage of patients who answered at least one of the questions 1 and 2 with "yes" under Suicidal Ideation
- Number and percentage of patients answering each of the questions 1 to 5 with "yes"
- For the patients who answered "yes" to questions 1 and 2:
  - Number and percentage in each category (1 to 5) for most severe suicide ideation
  - Answers (number and percentages per answer) to the Intensity of Ideation questions
- Number and percentages for the answers to the questions under Suicidal Behavior
- Number of actual suicide attempts

The same table – especially the number and frequency of patients who committed actual suicide attempts since the last visit – is used for the evaluation of C-SSRS as an exploratory efficacy endpoint.

Responses to each item will be listed by patient.

5.2.11 Safety Monitoring (Data Monitoring Committee [DMC])

A data monitoring committee (DMC) will be established to monitor the safety and efficacy data generated during the study. The DMC will consist of three independent clinical specialists. The DMC will review accumulated data from the ongoing study. Based on these reviews, the DMC will advise on the further conduct of the study. An independent unblinded off site statistician will present safety outputs to the DMC and conduct an interim analysis for futility or sample size re-calculation, see Section 4.8. The composition, activities, and responsibilities of the DMC is described separately in the DMC charter.

The following displays will be prepared for each regular DMC meeting:

- Disposition Table, as specified in the SAP, but also showing number of patients currently still on treatment
- Overall summary of Adverse Events
- Summary of Adverse Events by System Organ Class, Preferred Term and Treatment
- Summary of Compliance for Drug Intake
- Listing of Serious Adverse Events
- Listing of Adverse Events
- Listing of Abnormal Laboratory Values
- Listing of Columbia-Suicide Severity Rating Scale results

Details will be given in a separate IDMC TFL Shell document.

5.3 Other Analyses

The analysis of biomarker and genotyping data will be done at a later time and are not described in this SAP.

5.3.1 Pharmacokinetics

Pharmacokinetic concentrations will be listed by patient. The listing is based on the SAF, but patients excluded from the PPS will be flagged.

Summary statistics for the active treatment will be produced using the PPS.

Spaghetti plot and mean plots of concentrations versus visit will be displayed. The SAF will be used for individual concentrations and the PPS will be used for means.

When calculating means for figures and summary statistics values below the lower limit of quantification (BLQ values) will be replaced with 0.

For US Sites only, a sampling at around Tmax to assess average levels and help drug monitoring will be conducted in a patient subpopulation (approximately 10 patients). If Cmax values observed are in the range of those seen in healthy volunteers, no further samples should be necessary at Tmax. In this selected subpopulation, PK assessments will be obtained at Visits 4 and 6 (in addition to the pre-dose sample already scheduled on these visits as per the original protocol). This analysis will be described in detail in a separate SAP.

5.3.2 Pharmacodynamics

Target engagement (LSD-TE1) will follow the analysis for PK described in Section 5.3.1. In addition, LSD-TE1 is used in exposure analysis (see Section 5.2.1).

5.3.3 Visit to Health Care Services

The number of visits to primary care services, mental health services, emergency room services and the overall number of visits to any health care services in the month prior to the Visit will be summarized descriptively by Visit and treatment.

In addition, the number and frequency of patients for each reason of visit will be summarized. For post-baseline Visits, the change from baseline for the number of visits by type of service and reason will be calculated and included in the summary. The number and frequency of patients having visited health care services at least once will also be summarized by visit, type of service and reason for visit.

The total number of visits by type of services will be calculated for each patient across all Study Visits. As an exploratory analysis only, the Mann-Whitney test will be used to compare the total number between the two treatment arms. Only patients who do not discontinue early will be included in this analysis.

A listing by patient will be produced showing the number of visits and the reason for the visits.

5.3.4 Exploratory Assessments

The analysis of C-SSRS is described in Section 5.2.10, the analysis of visits to Health Care Services in Section 5.3.3 and the PK/PD analysis in Sections 5.3.1 and 5.3.2.

The Agitation-Aggression Psychiatric Inventory-Clinician Report (AAPI-CR) and Brief Assessment of Cognition (BAC) Scale results will be primarily analyzed using the following variables.

- AAPI-CR: Total Score, Agitation/Aggression Subscale and Psychiatric Subscale
- BAC: Standard-Composite TScore

Those variables will be listed and summarized descriptively by Visit and Time Point using the FAS. Additional listings will be produced with the results for each item.

5.4 COVID-19 Related Considerations

The impact of Covid-19 on the study will be investigated as follows:

- Covid-19 related discontinuations will be displayed in the disposition summary as a reason for discontinuation
- Number and frequency of Covid-19 related adverse events (i.e., infection) will be displayed on the main summary table for adverse events
- Number and frequency of concomitant medication for Covid-19 will be displayed on the main summary table for concomitant medication
- Covid-19 related protocol deviations will be summarized in an additional summary table
- For the primary efficacy analysis, patients who discontinue/could not complete the assessment due to Covid-19 will not be given special consideration and treated the same as any other patients with other reasons for missing data

5.5 Changes in the Conduct of the Study or Planned Analysis

There are no changes in the planned analysis compared to the study protocol.

6 REFERENCES

[1] Mehta CR, Pocock SJ. Adaptive Increase in Sample Size when Interim Results are Promising: A Practical Guide with Examples. *Statist. Med.* 2000;67:00:1-6.
[2] Schenker N., Taylor J.M.G. Partially Parametric Techniques for Multiple Imputation. *Computational Statistics & Data Analysis* 1996;Volume 22, Issue 4.
[3] Ratitch B, O'Kelly M, Tosiello R. "Missing Data in Clinical Trials: From Clinical Assumptions to Statistical Analysis Using Pattern Mixture Models." Pharm Stat. 2013;12:337-347.

Software:

[4] SAS® Version 9.2 of the SAS System for Personal Computers. Copyright © 2002-2003. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

7 Appendices

7.1 Schedule of Assessments

Table 1. Schedule of Assessments and Procedures

Study Procedures	Visit	V1 (Screening)	V2 (Baseline)	V3	V4	V5	V6	V7	V8 / EoS[2]	V9 Run-out (Safety F-up) [13]
	Week	**-1**	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**
	Day [1]	-7 to 1	(-5 to 1) ±5	15±2	29±2	43±2	57±2	71±2	85±2	99±2
Informed Consent Form [3]		X								
Inclusion/Exclusion Criteria		X	X							
Mini-International Neuropsychiatric Interview (MINI)		X								
Demographics		X								
Medical history		X								
Prior and concomitant medication		X	X	X	X	X	X	X	X	X
Physical examination (height –only at V1- and weight)		X	X	X	X	X	X	X	X	X
Tuberculosis status by medical history, signs and symptoms		X								
Urine pregnancy test (females of reproductive age)		X	X						X	
Urine drug screening		X	X	X	X	X	X	X	X	X
Subject Eligibility Form		X [4]								
Randomization			X							
Follow-up phone calls (only applicable if subject is receiving concomitant antidepressants) [12]		At V2 (Baseline): after 6-8h; 24h; and 48 h, after first drug administration								
Rating Scales/Clinical Outcomes Assessments (COAs)										
Columbia-Suicide Severity Rating Scale (C-SSRS)		X	X	X	X	X	X	X	X	X [5]
Agitation-Aggression Psychiatric Inventory-Clinician Report (AAPI-CR)		X	X		X		X		X	X [5]

Study Procedures	Visit	V1 (Screening)	V2 (Baseline)	V3	V4	V5	V6	V7	V8 / EoS[2]	V9 Run-out (Safety F-up) [13]
	Week	**-1**	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**
	Day [1]	-7 to 1	(-5 to 1) ±5	15±2	29±2	43±2	57±2	71±2	85±2	99±2
Brief Assessment of Cognition (BAC) Scale		X	X			X			X	X [5]
Beck Depression Inventory – II (BDI-II)		X	X	X	X	X	X	X	X	X [5]
Borderline Evaluation of Severity over Time (BEST)		X	X	X	X	X	X	X	X	X [5]
Borderline Personality Disorder Checklist (BPDCL)		X	X	X	X	X	X	X	X	X [5]
State-Trait Anxiety Inventory (STAI) [6]		X	X	X	X	X	X	X	X	X [5]
State-Trait Anger Expression Inventory 2 (STAXI-2)		X	X	X	X	X	X	X	X	X [5]
Clinical Global Impression-Severity for Agitation/Aggression (CGI-S A/A)		X	X	X	X	X	X	X	X	X [5]
Number of visits to healthcare services outside of the trial (primary care, mental health services & ER visits)			X [7]		X [7]		X [7]		X [7]	X [5]
Safety Assessments										
Adverse events			X	X	X	X	X	X	X	X
Vital signs [8]		X	X	X	X	X	X	X	X	X
Electrocardiogram [8]		X							X	
Blood sample for laboratory safety assessments [8]										
Hematology		X	X	X	X	X	X	X	X	X [5]
Biochemistry		X	X	X	X	X	X	X	X	X [5]
Hep C, Hep B, HIV testing		X								
Blood sample for pharmacokinetic and target engagement [9]										
PK [9i]			X		X[9i]		X[9i]		X	
LSD1-TE			X		X		X		X	
Blood sample for exploratory assessments [10]										
Biomarker's analysis			X						X	

Study Procedures	Visit	V1 (Screening)	V2 (Baseline)	V3	V4	V5	V6	V7	V8 / EoS[2]	V9 Run-out (Safety F-up)[13]
	Week	**-1**	**0**	**2**	**4**	**6**	**8**	**10**	**12**	**14**
	Day[1]	-7 to 1	(-5 to 1) ±5	15±2	29±2	43±2	57±2	71±2	85±2	99±2
Genotyping			X							
Investigational Medicinal product (IMP)										
Assess IMP compliance (manual accountability of returned IMP at site)				X	X	X	X	X	X	X[5]
Dispense IMP[11]			X	X	X	X	X	X	X[5]	

** Screening period will last a week with a window of ± 5 days.*

(1) All assessments from a specific visit may be completed over a maximum of two consecutive days. In this case, the first day should be considered as the visit day. The forecasting for visit schedule should take V2 (Baseline) date as the reference. If the date of a visit changes within the allowed window, the following visits will still be scheduled based on V2 (Baseline) date.

(2) After Visit 2, seven visits, every two weeks, are planned until Visit 9. If subjects do not complete all study visits (early termination), then they should complete all the Visit 8/EoS assessments. All subjects are classified as follows:

a. *Completers:* Subjects completing Visit 9.

b. *Non-completers:* Subjects that received at least one dose of study treatment, but prematurely discontinued the study before completion of Visit 9 for any reason other than withdrawal of the informed consent. These subjects will be asked to complete an End of Study (EoS) Visit as soon as possible after withdrawal. Non-completers will not enter the 2-week subject-blind placebo run-out period.

   i. If there has been, at least, one week between the last intake of study drug and this EoS visit, no further Safety follow-up visit will be required.

   ii. If there has been less than one week between the last intake of study drug and EoS visit, subject will be asked to attend to a Safety-follow up visit, two weeks after the last intake of study drug, to perform safety assessments only.

c. *Non-completers who discontinue the study because they withdraw their consent:* Subjects that received at least one dose of study treatment, but prematurely discontinued from the study before completion of Visit 9 due to withdrawal of informed consent. These subjects, if possible, will be asked to attend to an EoS Visit.

   i. The visit must be scheduled as soon as possible after withdrawal.

   ii. If the subject withdraws consent during a visit but agrees to complete the visit, then, the investigator will complete an EoS Visit. All the data collected up to and including this visit will be used in the analysis.

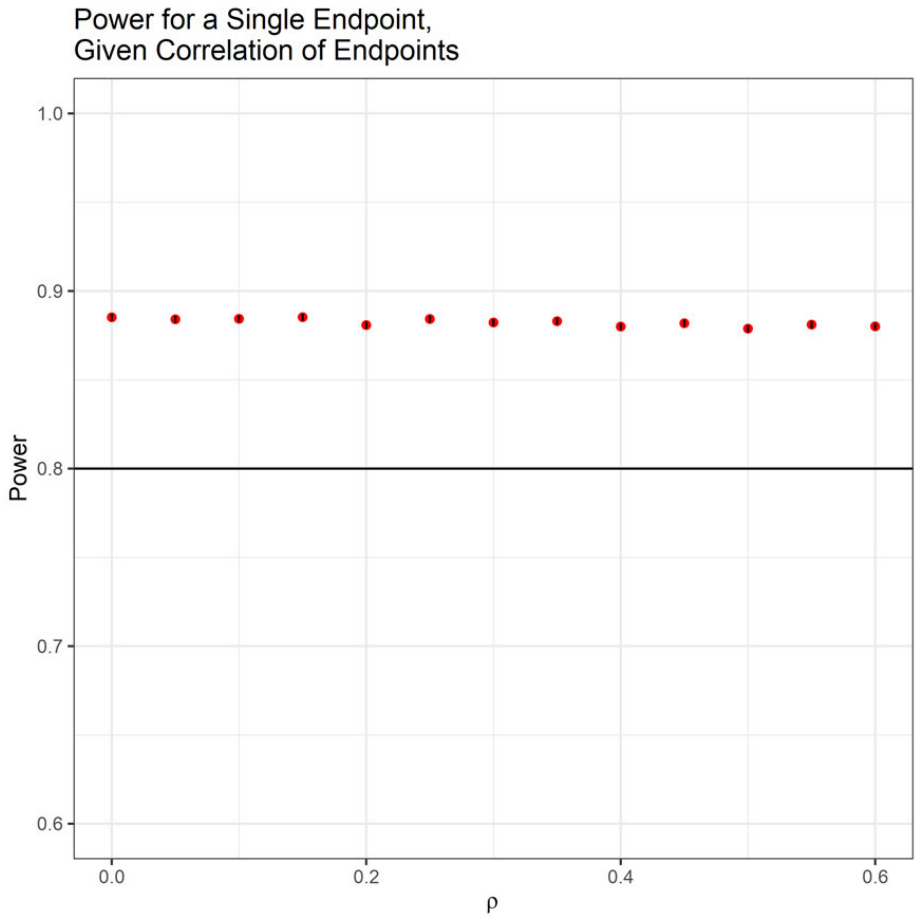
   iii. If the subject withdraws consent and refuses to complete the EoS Visit, no new information will be collected. However, it is recommended,

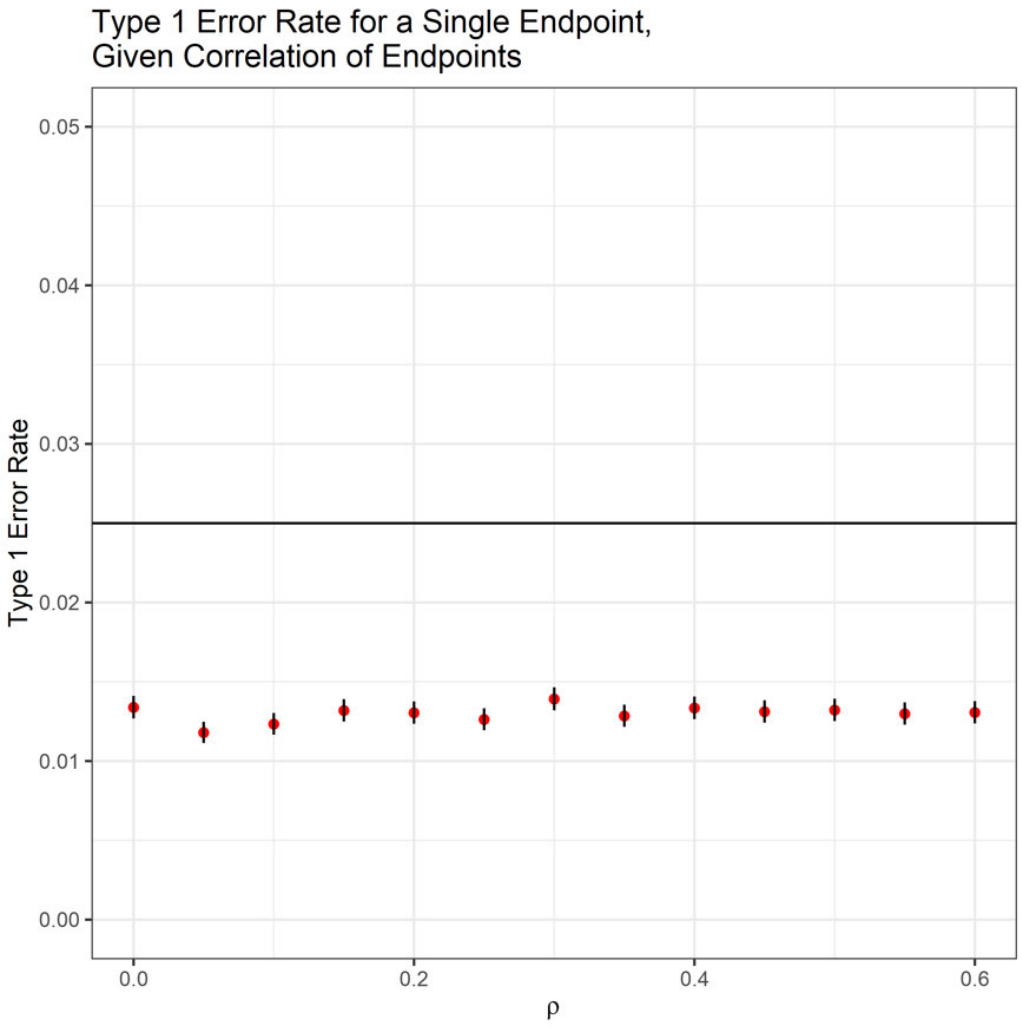
outside the scope of this study, to perform a safety assessment within 2 weeks after treatment discontinuation.

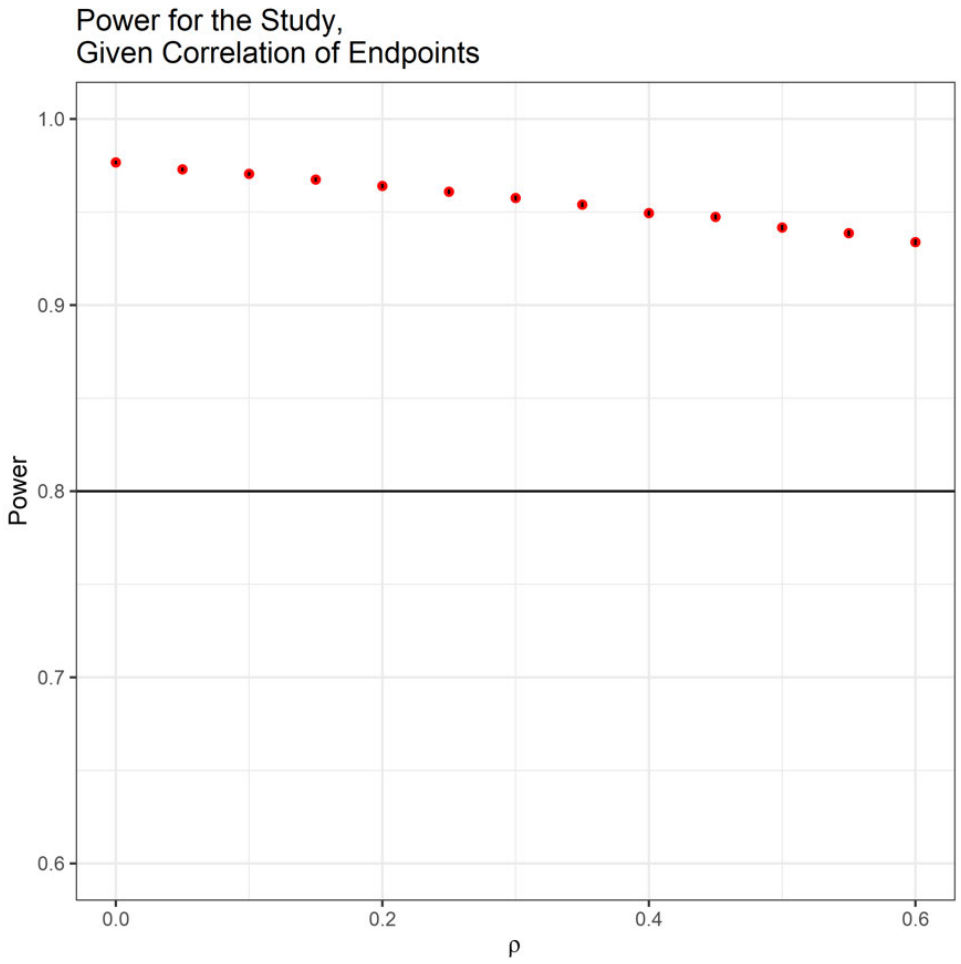
(3) Informed consent at the Screening visit must be obtained before any study specific procedure.
(4) Participant eligibility should be documented in a Subject Eligibility Form and submitted to the study medical monitor immediately after Visit 1 for approval before Visit 2.
(5) Not applicable for non-completers (i.e.: subjects who received at least one dose of study treatment but prematurely discontinue the study before its completion).
(6) The full State-Trait Anxiety Inventory (STAI) (Forms Y-1 & Y-2) will only be administered at Screening (Visit 1), Baseline (Visit 2) and Visit 8/EoS, whereas Form Y-1 ("state" anxiety) will be administered at all other visits.
(7) The number of visits to healthcare services outside of the trial (primary care, mental health services & ER visits) will be evaluated 'in the last month' at Baseline and at Visit 4, Visit 6 and Visit 8/EoS.
(8) Vital signs and electrocardiogram (ECG) should be obtained for a subject in the same manner throughout the study (e.g., obtained from the same arm).
(9) Blood extractions should take place at the same time of the day and in fasting conditions (first thing in the morning preferably, in order to allow subjects to ingest food before going on with the rest of the study assessments). In case the study visit is scheduled in the afternoon, fasting conditions should last 6-8 hours. Subjects should be instructed to take IMP in fasting conditions.
(9i) For US Sites ONLY: Cmax will be assessed in a subset of subjects. In these subjects, one extra blood sample for PK assessments will be obtained 1 to 2 hours after IMP intake on Visits 4 and 6 (in addition to the pre-dose sample already scheduled on these visits). Subjects will have to sign a separate specific informed consent form.
(10) As blood sampling for the exploratory assessments is an integral part of the study, the main Patient Information Sheet covers these analyses
(11) Study treatment will be dispensed at each study visit from Visit 2 to Visit 8. Each dispensation will consist in a box containing the treatment for 14±2 days. At each visit, the first study drug intake, corresponding to the new treatment box dispensed on that visit, should occur at the clinic.
(12) ONLY APPLICABLE IN CASE THE SUBJECT IS RECEIVING CONCOMITANT TREATMENT WITH ANTIDEPRESSANTS (i.e., selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, tricyclics, tetracyclics, and triazolopyridines): A follow-up call will be required approximately 6-8 hours post first IMP administration, as well as at 24 hours and 48 hours to ensure subjects are not experiencing any signs or symptoms of serotonin syndrome. If a subject is experiencing moderate symptoms of serotonin toxicity such as agitation, tachycardia, diaphoresis, and/or hyperthermia, then they will immediately be referred to local emergency services.
(13) For US Sites ONLY: A subset (i.e., 3-4) of subjects and the PI at each US Site will complete qualitative research regarding their experience with some of the measures used in PORTICO (i.e., BPDCL for subjects & CGI-S A/A and AAPI-CR for clinicians) at Visit 9. Subjects will have to sign a separate specific informed consent form.

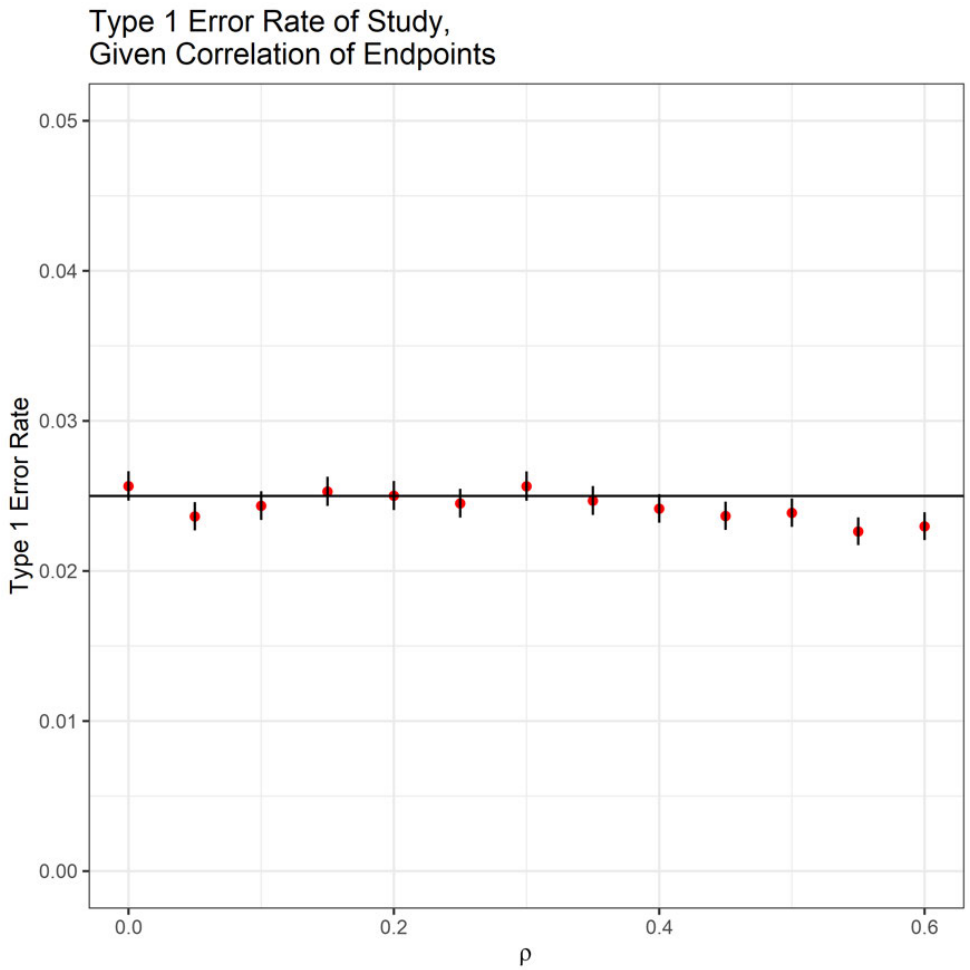
7.2 Promising Zone Type 1 Error Rate Simulation

7.2.1 Figure Output




Type 1 Error Rate for a Single Endpoint,
Given Correlation of Endpoints
Type 1 Error Rate
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.2
0.4
0.6
ρ


Power for the Study,
Given Correlation of Endpoints
Power
1.0
0.9
0.8
0.7
0.6
0.0
0.2
0.4
0.6
ρ


Type 1 Error Rate of Study,
Given Correlation of Endpoints
Type 1 Error Rate
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.2
0.4
0.6
ρ

7.3 Promising Zone Decision Tree

- If estimates from both outcomes are in the promising zone and the difference between the suggested sample sizes for completers is greater than 60 individuals then the least conservative sample size (smallest completer sample size) will be reported. Otherwise, the most conservative (largest completer sample size) will be reported.

- If the estimate from one outcome is in the promising zone and the other in the favorable zone and the difference between the suggested sample sizes for completers is greater than 60 individuals then the least conservative sample size (smallest completer sample size/favorable zone (150)) will be reported. Otherwise, the most conservative (largest completer/promising zone sample size) will be reported.

- If the estimate for one outcome falls in the unfavorable zone and one in the promising zone, then the promising zone sample size increase will be reported.

- If the estimate one outcome falls in the promising zone and one in futility, then the promising zone sample size increase will be proposed.

- If the estimates for both outcomes fall in the unfavorable zone, then there is no change in the sample size.

- If the estimate for one outcome falls in the favorable and the other unfavorable zone, then there will be no suggested change in sample size.

- If the estimate one outcome falls below the threshold for futility and the estimate for the other outcome falls in the unfavorable zone, then there will be no suggested change in sample size.

- If the estimates for both outcomes fall in the favorable zone, then there will be no suggested change in sample size.

- If the estimate for one outcome falls below the threshold for futility and the other is in the favorable zone, then there is no change in the sample size.

- If the estimates for both outcomes are below the threshold for futility, then the algorithm will recommend stopping for futility.